Stratification Directionnelle adaptative

Miguel Munoz Zuniga, Josselin Garnier, Emmanuel Remy, Etienne de Rocquigny

Stratification Directionnelle Adaptative

Miguel Munoz Zuniga[1,2], Josselin Garnier[1], Emmanuel Remy[2] & E. de Rocquigny[2]

[1] *Laboratoire de probabilités et modèles aléatoires, Université Paris VII, Case courrier 7012, 2 place Jussieu, 75251 Paris cedex 05, France*

[2] *EDF R&D, 6 quai Watier, 78400 Chatou, France*

Résumé

Nous présentons dans cet article une nouvelle méthode de Monte-Carlo accélérée, nommée SDA : Stratification Directionnelle Adaptative, developpée pour l'estimation de faibles probabilitées avec un nombre limité de simulations. Celle-ci couple la méthode de stratification, permettant la mise en place d'une méthode adaptative à tirages préférentiels, et la simulation directionnelle, adaptée à l'estimation de faibles probabilités et possédant un bon rapport "précision/temps de calculs". Nous proposons tout d'abord un estimateur avec deux étapes, apprentissage puis estimation, SDA-2, généralisé ensuite en une méthode à $L-1$ étapes d'apprentissage et une étape d'estimation, SDA-L. Dans les deux cas, nous étudions les comportements asymptotique et non asymptotique des estimateurs. Nous proposons également une amélioration de SDA-2, pour solutionner le problème d'espace d'exploration trop vaste engendré par une dimension trop élevée. Pour cela, nous introduisons une statistique, évaluée à la fin de l'étape d'apprentissage, permettant une réduction pertinente du nombre de strates et une classification des variables aléatoires selon leur influence sur la défaillance.

Mots clés : Fiabilité, Probabilité de défaillance, Simulation, Directionnelle, Stratification

Abstract

The aim of this paper is to present a new accelerated Monte-Carlo simulation method, named ADS: Adaptive Directional Stratification, designed to estimate small probabilities with a limited number of simulations. This new stochastic technique is an original variant of adaptive accelerated simulation method, combining stratified sampling, which enables to achieve an adaptive importance sampling method, and directional simulation, well adapted to small failure probability estimation and giving a good "precision/calculation time" ratio. We propose an estimator with two steps, learning and then estimation, ADS-2, and we generalize it with a scheme with $L-1$ learning steps, ADS-L. First, we theoretically study the asymptotic and non-asymtotic properties of the estimators. Then, we propose some improvements for our new method. To begin with, to overcome the limit involved by high dimensional inputs, we introduce the ADS-2^+ method, which has the same ground as the ADS-2 method, but additionnally uses a statistical test to detect the most significant inputs, and carries out the stratification only along them.

Keywords: Reliability, Failure probability, Directional, Sampling, Stratification

1 Introduction

De manière générale, l'objectif de ce travail est l'estimation de l'espérance $I = \mathbb{E}\big(F(\mathbf{X})\big)$ où $\mathbf{X}$ est un vecteur aléatoire p-dimensionnel défini sur l'espace probabilisé $(\mathbb{D}, \mathcal{A}, \mathbb{P})$ et F une fonction mesurable bornée. Dans le cadre de la fiabilité des structures, on choisit $F(\mathbf{x}) = \mathbb{1}_{G(\mathbf{x})<0}$ avec $G : \mathbb{D} \subset \mathbb{R}^p \to \mathbb{R}$ et l'espérance à estimer devient une probabilité de défaillance : $I = \mathbb{P}\big(G(\mathbf{x}) < 0\big)$. Nos contraintes sont alors les suivantes :

(C_0) : G est complexe et coûteuse en ressources computationnelles et par conséquent, nous sommes limités à un maximum de quelques milliers d'appels à G,

(C_1) : la défaillance est un évènement rare, i.e. I est faible. Nous considérerons qu'une probabilité est faible si elle est inférieure à 10^{-3}. En effet, pour des valeurs de probabilités plus élevées, il existe déjà de nombreuses méthodes efficaces (Monte-Carlo standard, FORM/SORM, tirage d'importance...),

(C_2) : le résultat doit être robuste, i.e. avec un contrôle sur l'erreur de l'estimation. Ce résultat peut en effet être utilisé dans un processus de décision (programme de maintenance-inspection, décision d'extension de durée de vie...) et doit donc s'accompagner d'un minimum de garanties.

2 La méthode SDA

L'idée est de coupler la stratification et la simulation directionnelle en une seule méthode, afin d'exploiter les possibilités offertes par ces dernières : stratégie adaptative, tirages préférentiels, résultats d'allocations optimales, estimation efficace de faibles probabilités et temps de calculs raisonnables. La méthode se décompose en une phase préliminaire de reformulation du problème (Pr), une phase d'apprentissage (Ap) et une phase d'estimation (E). La phase préliminaire se décompose en 4 points.

(Pr_1) Transformer le vecteur aléatoire $\mathbf{X}$ en un vecteur aléatoire $\mathbf{U}$ à composantes Gaussiennes centrées réduites et indépendantes. Pour cela, il existe diverses transformations : nous utiliserons celle de Nataf, voir Nelsen (1999).

(Pr_2) Décomposer $\mathbf{U}$ en coordonnées polaires, i.e. en un angle aléatoire $\mathbf{A}$ suivant une loi uniforme sur la sphère unité, $Unif(S_p)$, et un rayon aléatoire R, tel que R^2 suit une loi du Chi-2 à p degrés de liberté. Cette décomposition est toujours possible si $\mathbf{U}$ suit une loi sphérique, ce qui est bien le cas, voir Nelsen (1999).

(Pr_3) Décomposer I conditionnellement à $\mathbf{A}$.

(Pr_4) Stratifier l'espace aléatoire des directions, autrement dit la sphère unité, en "quadrants", voir Munoz Zuniga et al. (2009). En effet, les cônes sont des formes de

strates naturellement adaptées à un sondage directionnel de l'espace et le fait que nous supposions n'avoir aucune information a priori sur l'espace de recherche motive le choix de quadrants.

Nous noterons n le nombre total de simulations directionnelles, i.e. le nombre total de tirages du vecteur aléatoire $\mathbf{A}$. Pour mettre en oeuvre les 2 phases d'apprentissage et d'estimation, nous avons besoin de décomposer n en deux parties. Soit $\gamma_1(n)$ le pourcentage de simulations alloué à la phase d'apprentissage et $\gamma_2(n) = 1 - \gamma_1(n)$ celui alloué à la phase d'estimation. La phase d'apprentissage se décompose alors en 4 points, voir Munoz Zuniga et al. (2009).

(Ap_1) Déterminer une allocation a priori du nombre $\gamma_1(n)n$ de simulations directionnelles par quadrant. Sans aucune information préalable, nous choisirons une allocation uniforme.

(Ap_2) Estimer la probabilité de défaillance dans chaque direction simulée, par un algorithme déterministe de recherche de zéros requérant un certain nombre d'évaluations à G et en déduire une estimation de la probabilité de défaillance par quadrant.

(Ap_3) Estimer dans chaque quadrant la variance de l'estimation de la probabilité.

(Ap_4) Estimer l'allocation optimale grâce aux précédentes estimations de variance par quadrant. La formule d'estimation de l'allocation optimale dépendra du choix de réutiliser ou non les résultats des simulations de la phase d'apprentissage, autrement dit du choix de considérer un estimateur avec ou sans recyclage.

Finalement, la phase d'estimation se décompose en 3 étapes, voir Munoz Zuniga et al. (2009).

(E_1) Simuler les $\gamma_2(n)n$ tirages directionnels selon l'allocation optimale estimée en (Ap_4).

(E_2) Estimer la probabilité de défaillance, dans chacune des nouvelles directions simulées, par un algorithme déterministe de recherche de zéros.

(E_3) Estimer I, soit par un estimateur avec recyclage, $\hat{I}_r^{SDA-2}$, utilisant les n simulations, soit par un estimateur sans recyclage, $\hat{I}_{nr}^{SDA-2}$, utilisant uniquement les $\gamma_2(n)n$ dernières simulations.

En résumé, pour réaliser cette méthode, nous avons théoriquement : (a) déterminé les allocations optimales à réaliser dans chacun des quadrants dans les cas avec et sans recyclage, (b) démontré que l'estimateur sans recyclage est non biaisé et que l'estimateur avec recyclage est biaisé avec un biais difficilement estimable sous la contrainte (C_0), donc délicat d'utilisation en pratique, (c) déterminé dans un cadre non asymptotique l'expression des variances des deux estimateurs, ce qui a fourni des formules théoriques utilisables pour

l'estimation de la variance de l'estimateur sous la contrainte (C_0). Enfin, dans un cadre asymptotique, nous avons démontré la convergence des deux estimateurs sous certaines hypothèses sur l'allocation du nombre de simulations entre les phases d'apprentissage et d'estimation. Premièrement, si nous supposons que $\gamma_1(n) \to 0$ et $\gamma_1(n)n \to +\infty$, alors :

$$\sqrt{n}(\hat{I}_r^{SDA-2} - I) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2_{opt_1}) \quad \text{et} \quad \sqrt{n}(\hat{I}_{nr}^{SDA-2} - I) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2_{opt_1}). \tag{1}$$

Deuxièment, si nous supposons que $\gamma_1(n) = \gamma_1$, alors :

$$\sqrt{n}(\hat{I}_r^{SDA-2} - I) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2_{opt_2}) \quad \text{et} \quad \sqrt{n}(\hat{I}_{nr}^{SDA-2} - I) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \tfrac{1}{1-\gamma_1}\sigma^2_{opt_1}) \tag{2}$$

avec $\sigma^2_{opt_1}$ et $\sigma^2_{opt_2}$ les variances minimales obtenues avec les réelles allocations optimales, qui de plus vérifient : $\sigma^2_{opt_1} \leq \sigma^2_{opt_2}$.
On remarquera que, dans notre méthode, nous n'adaptons pas les strates, les méthodes et outils nécessaires étant trop coûteux en termes de nombre d'appels à G. Pour illustrer ce point, on pourra consulter Etoré et al. (2009), qui propose une méthode de stratification adaptative, avec adaptation des strates, basée sur Cannamela et al. (2008) et Etoré et al. (2008).
Nous sommes capables de généraliser la méthode SDA-2 en une méthode multi-adaptative. Deux points de vue sont possibles. Le premier consiste à se donner un nombre d'étapes $L \geq 2$ fixé, de telle sorte que l'on effectuera $L-1$ étapes d'apprentissage et une étape d'estimation , puis à construire notre estimateur, $\hat{I}^{SDA-L}$, et à étudier finalement son comportement asymptotique lorsque n deviendra grand. Le second consiste, pour un nombre d'étapes L, à construire un estimateur puis à regarder son comportement asymptotique lorsque L et n deviendront grands : on notera cet estimateur $\hat{I}^{SDA-\infty}$. Sous certaines hypothèses présentées dans Munoz Zuniga et al. (2009), les résultats asymptotiques escomptés sont bien obtenus. Cependant, bien que la méthode multi-adaptative soit attirante théoriquement, il semble clair qu'elle ne pourra pas être mise en oeuvre sous la contrainte (C_0).

3 Fléau de la dimension : SDA-2^+

Lorsque la dimension du vecteur aleatoire $\mathbf{X}$ augmente, le nombre de strates augmente exponentiellement : en effet, en dimension p, le nombre de quadrants est de 2^p. Comme un minimum de simulations est nécessaire pour explorer chaque quadrant, le nombre de simulations directionnelles devient vite trop grand. Sous C_0, la méthode SDA-2 reste efficace pour une dimension inférieure ou égale à 4. Pour pouvoir considérer des problèmes de dimensions supérieures, l'idée est d'obtenir, à la fin de la phase d'apprentissage, un classement des composantes de $\mathbf{X}$ en fonction de leur influence sur la défaillance, pour ensuite, lors de l'étape d'estimation, stratifier uniquement selon les plus importantes.

Pour déterminer si une composante aléatoire X_k sera stratifiée, pour $k \in \{1, ..., p\}$, nous proposons la méthode suivante. Tout d'abord, nous indexons chaque quadrant. Pour tout $k \in \{1, ..., p\}$, on donne à la k-ème composante de $\mathbf{X}$ une étiquette i_k à valeur dans $\{-1, 1\}$ correspondante au signe de X_k. Ainsi, chaque quadrant est caractérisé par un p-uplet $(i_1, ..., i_p)$. Puis, nous définissons la suite $T := (T_k)_{k=1,...,p}$ telle que :

$$T_k = \sum_{i_l \in \{-1,1\}, l \neq k} \left| \tilde{I}(i_1, ..., i_{k-1}, -1, i_{k+1}, ..., i_p) - \tilde{I}(i_1, ..., i_{k-1}, 1, i_{k+1}, ..., i_p) \right| \qquad (3)$$

pour tout $k \in \{1, ..., p\}$, avec $\tilde{I}(i_1, ..., i_p)$ l'estimation de la probabilité de défaillance dans la strate $(i_1, .., i_p)$ obtenue lors de la phase d'apprentissage de la méthode SDA-2. Chaque T_k agrège les différences de probabilité de défaillance entre les quadrants le long de la dimension k. Plus T_k est grand, plus l'influence de la k-ème composante sera influente. Donc nous proposons de classer la suite T par ordre décroissant et de stratifier uniquement les $p' < p$ premières composantes, les autres composantes étant alors simulées sans stratification. Une étude numérique poussée a permis de constater que la méthode SDA-2 est efficace pour une dimension autour de $p = 3$. Par conséquent, nous avons opté pour un nombre de composantes stratifiées $p' = 3$. Une fois celles-ci déterminées, nous recalculons les allocations optimales à effectuer dans les nouveaux $2^{p'}$ "hyper-quadrants" et achevons la méthode en réalisant la phase d'estimation avec cette nouvelle allocation estimée. Nous appellerons cette version optimisée SDA-2^+. Sous la contrainte (C_0), la méthode SDA-2^+ permet de considérer des vecteurs aléatoires de dimension 5 et 6. Pour des dimensions supérieures ou égales à 7, l'espace à sonder devient trop vaste pour espérer obtenir des estimations correctes des allocations optimales par strate. Dans ce cas, des études préliminaires visant à réduire le nombre de composantes du vecteur $\mathbf{X}$ sont fortement conseillées et réalistes d'un point de vue industriel. Une alternative est de fixer les $p - p'$ composantes non stratifées à des valeurs déterministes et conservatives et de réaliser la phase d'estimation dans l'espace de dimension p', pour ainsi obtenir une estimation et un demi-intervalle de confiance de la probabilité de défaillance. Ce genre de procédure est courante en analyse de risques, mais une mauvaise classification des composantes via T entraînera une estimation souvent trop conservative pour être d'une quelconque utilité. Une dernière solution est de continuer à simuler les $p-p'$ composantes non stratifiées par un Monte-Carlo standard : mais cette stratégie réduit considérablement la précision des estimations lorque le nombre de variables influentes est supérieur à p'.

4 Conclusion et perspectives

Pour résoudre le problème de l'estimation d'une probabilité de défaillance sous les contraintes (C_0), (C_1) et (C_2), nous avons développé une nouvelle méthode de Monte-Carlo accélérée SDA-2. Pour une dimension du vecteur aléatoire $\mathbf{X}$ inférieure ou égale à 4, SDA-2 répond correctement aux attentes. Quand la dimension est supérieure ou égale à

5, comme le nombre de strates croît exponentiellement et que l'espace à sonder devient trop vaste, SDA-2 ne donne plus des résultats suffisamment robustes. Alors, pour contrer le "fléau de la dimension", nous avons développé SDA-2^+, intégrant une statistique pour la détection des variables significatives et nous permettant ainsi de réduire le nombre p de dimensions le long desquelles est appliquée la stratification. Les résultats obtenus restent alors robustes jusqu'à une dimension inférieure ou égale à 6. En quelque sorte, l'inverse de la distance entre p' et le vrai nombre de composantes influentes pondère l'amélioration apportée par SDA-2^+. Enfin, pour des dimensions au-delà de 6, des méthodes conservatives restent une solution possible. Il est à noter que nous avons appliqué la méthode SDA, d'une part à des exemples académiques, afin de calibrer au mieux certains de ses paramètres, et d'autre part à des cas industriels (via la plateforme OpenTURNS) pour lesquels des résultats performants ont été obtenus comparativement aux méthodes mises en oeuvre jusqu'à présent.
Parmi les perspectives, nous imaginons tout d'abord qu'une étude plus précise de la statistique T permettrait d'estimer plus efficacement le nombre de variables influentes. Un autre travail est envisageable sur la phase d'apprentissage, afin de rendre plus exacte la détection des strates importantes. Premièrement, afin de mieux sonder l'espace, une approche par méthode de quasi Monte-Carlo directionnelle, pour la phase d'apprentissage, pourrait s'avérer efficace. Deuxièmement, une étude sur la variation des allocations par strates estimées, via une méthode de bootstrap, permettrait de donner un critère de passage ou non à la phase d'estimation suite à la phase d'apprentissage. Finalement, des méthodes hybrides couplant SDA et des méthodes d'approximation numérique type FORM/SORM sont aussi imaginables, ces dernières offrant des informations pertinentes avec un nombre de simulations faible.

Remerciements : Ce travail a été partiellement financé par l'Agence Nationale de la Recherche dans le cadre du projet ANR OPUS (Open source Platform for Uncertainty treatment in Simulation), réf. ANR-07- CIS7-010 et ANR-07-TLOG-015.

Bibliographie

[1] Nelsen, R.B. (1999) An introduction to copulas. Springer.
[2] Munoz Zuniga, M., Garnier, J., Remy, E., de Rocquigny, E. (2009) Adaptive Directionnal Stratification - An adaptive directional sampling method on a stratified space. *Communication at ICOSSAR congress.*
[3] Cannamela, C., Iooss, B., Garnier, J. (2008) Controlled stratification for quantile estimation. *Annals of Applied Statistics*
[4] Étoré, P., Jourdain, B. (2008) Adaptive Optimal Allocation in stratified Sampling Methods. *Methodol. Comput. Appl. Probab.* (online first)
[5] Étoré, P., Fort, G., Jourdain, B., Moulines, É. (2009) On adaptive stratification. *Annals of Operations Research.* (online first)